Biochemistry and genetics of mannan-binding lectin (MBL)

J.S. Presanis, M. Kojima and R.B. Sim[1]
MRC Immunochemistry Unit, Department of Biochemistry, Oxford University, South Parks Road, Oxford OX1 3QU, U.K.

Abstract

Mannose- or mannan-binding lectin (MBL) is a member of the collectin protein family, which includes lung surfactant proteins SP-A and SP-D. Each member consists of similar or identical polypeptide chains with a region of collagen-like sequence followed by a C-type lectin domain. The polypeptides associate in threes to form a subunit containing a collagen-like helix, with three clustered lectin domains. These subunits associate into larger structures, usually with 12–18 polypeptides. The collectins bind to patterns of neutral sugars on surfaces (e.g. of micro-organisms) and mediate effector functions associated with killing/phagocytosis. MBL is the only collectin which activates complement. It resembles in quaternary structure the complement protein C1q, which recognizes targets via charge clusters. Binding of MBL to a surface activates MBL-associated serine proteases (MASPs) attached to MBL, and MASP-2 activates complement proteins C4 and C2. The MASPs are homologous to the C1q-associated proteases, C1r and C1s. MBL therefore activates complement by a mechanism very similar to C1q, and engages the opsonic activity of complement to clear micro-organisms. The serum concentration of MBL is very variable in humans. The variability is largely associated with mutations leading to amino acid substitutions in the collagen-like region which decrease MBL assembly and stability. Many studies demonstrate that MBL deficiency is associated with susceptibility to a range of infectious and inflammatory diseases.

MBL structure

Mannan-binding lectin (MBL) is a plasma glycoprotein of the collectin family [1], members of which bind to carbohydrates on the surfaces of micro-organisms and particulate materials (including altered host material, such as apoptotic cells), via their C-terminal lectin domains. It is also known as mannan-binding protein (MBP) or mannose-binding lectin, since it binds to yeast mannan [2] as well as to mannose coupled to Sepharose [3]. Previous names for MBL include core-specific lectin (CSL), from its interaction with the core motif of N-linked oligosaccharide [4], and Ra reactive factor (RaRF), as it binds to Ra chemotype strains of *Salmonella* [5]. The term MBL is now used in preference to MBP to avoid confusion with identically abbreviated proteins (e.g. major basic protein and myelin basic protein) and to emphasize its lectin character.

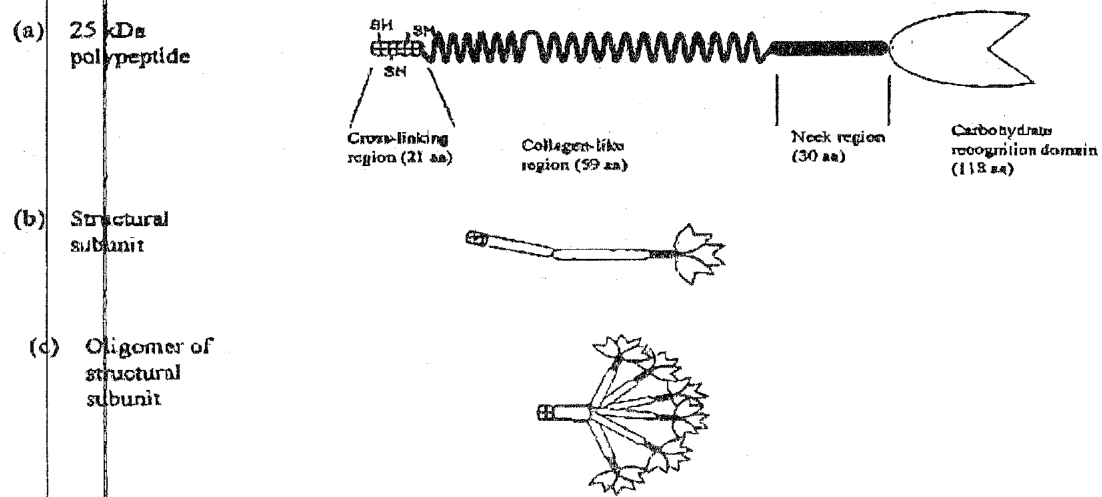
The polypeptide chain of secreted MBL is 228 amino acids long (not including the 20-residue signal peptide), and consists of a 20-residue 'cysteine-rich' region (containing 3 cysteines), followed by a collagenous region containing 19 Gly-Xaa$_1$-Xaa$_2$ triplets, a 'neck' region and then a C-terminal calcium-dependent carbohydrate-binding lectin domain (also called a carbohydrate-recognition domain (CRD) [6–9]). The neck region forms an α-helical coiled-coil structure which probably promotes trimerization of three polypeptides to form the subunit. The trimer is stabilized by hydrophobic interactions and inter-chain disulphide bonds within the N-terminal cysteine-rich region. MBL subunits assemble into larger oligomeric structures forming a 'bunch-of-tulips' or sertiform appearance (Figure 1) [10]. The commonest oligomeric form in humans appears to be a six-subunit (18-polypeptide chain) form with an overall molecular mass of about 18 × 25 000 Da. Unusually, disulphide bridging within and between subunits is incomplete and variable, so that, for example, the common six-subunit form is heterogeneous, consisting of a number of isoforms with different disulphide bridging. This is evident from SDS/PAGE analysis of non-reduced MBL, and comparison with analyses made by non-denaturing hydrodynamic methods [11,12]. Several publications (e.g. [13,14]) suggest that native MBL in serum or plasma occurs in oligomeric forms of different sizes (ranging from one to six subunits in humans), but it appears likely that the six-subunit oligomer is by far the major form [15]; smaller oligomers may form on storage or processing of plasma. Other proteins with similar quaternary structure also show variable polymerization: C1q, for example, has a six-head and a minor two-head form; the collectin SP-A (surfactant protein-A) also occurs in forms with one to six subunits [16]. The oligomerization of MBL allows avid binding to carbohydrate ligands due to the multiple CRDs present; forms with a lower degree of polymerization bind less avidly to carbohydrate surfaces, and are also defective in activating complement [12].

MBL has no N-linked carbohydrate, but its collagenous region contains hydroxylated lysine and proline residues, which have O-linked glycosylation.

Key words: complement, immunity, mannan-binding lectin (MBL), protease.
Abbreviations used: CRD, carbohydrate-recognition domain; MBL, mannan-binding lectin; MASP, MBL-associated serine protease; PAMP, pathogen-associated molecular pattern.
[1]To whom correspondence should be addressed (e-mail bob.sim@bioch.ox.ac.uk).

Figure 1 | Structure and organization of MBL

(**a**) MBL consists of an N-terminal cross-linking region, a collagen-like region, a neck region and a CRD. These 25-kDa polypeptides trimerize to form a structural subunit (**b**) that is found in serum as oligomers (**c**) [26].



MBL-binding specificity

MBL requires calcium ions for binding to its carbohydrate ligands. A calcium ion directly participates in ligand binding, forming co-ordination bonds to the 3- and 4-OH groups of the target sugar. Weis et al. [17] demonstrated this and the requirement for the 3- and 4-OH groups to be in the equatorial plane of the sugar ring structure. This requirement is due to hydrogen bonding to amino-acid side chains and restricts the binding of MBL to *N*-acetyl-D-glucosamine, mannose, *N*-acetylmannosamine, L-fucose and glucose [10]. The repeating sugar structures on microbial surfaces, not generally found on mammalian surfaces, can bind with high avidity to the CRDs which, within the trimeric head of each subunit, are spaced 45–55 Å apart. The specificity of MBL in recognizing sugar patterns on surfaces (pathogen-associated molecular patterns or PAMPs) relies therefore on the identity of the monosaccharide, its exposure on the surface and spacing between sugar residues. The spacing between the individual CRDs within one subunit, and the spacing of subunit heads relative to each other, are the determining factors (for further discussion see [18]).

Various studies have shown that MBL binds to clinically relevant micro-organisms including HIV-1 [19], influenza virus [18], *Staphylococcus aureus* [20], *Neisseria meningitidis* [21], *Candida albicans* [22] and *Leishmania major* [23].

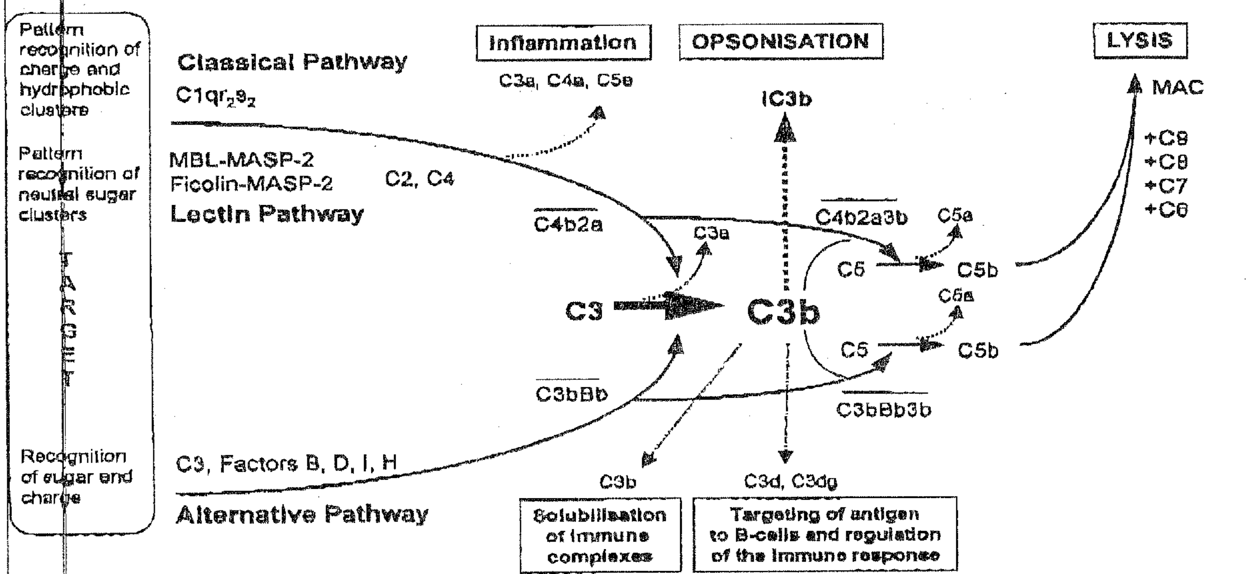
MBL serum levels

MBL serum levels are extremely variable, ranging from almost 0 to >5 μg/ml in healthy humans. MBL deficiency is common (5% or more of the population, depending on the threshold concentration used to define deficiency) and is associated with severe and repeated infections in infants. Soothill and Harvey [24] described a group of infants with recurrent pyogenic infections, whose serum failed to opsonize *Saccharomyces cerevisiae*. A similar opsonic defect was found in about 5% of an apparently healthy adult control population. Super and colleagues [25] linked this common opsonization deficiency with low MBL expression.

The variation in MBL levels can be attributed to three structural mutations of the MBL gene which are likely to result in defective polymerization, interacting with several polymorphisms in the MBL promoter region which influence the level of expression. The structural mutations occur at high frequency (generally 15% or greater cumulative allele frequency in most populations studied) and are single base changes in codons 52, 54 and 57 of exon 1. The changes are as follows: Arg-52 to Cys (R52C, MBL D variant), Gly-54 to Asp (G54D, MBL B variant) and Gly-57 to Glu (G57E, MBL C variant). The A variant is wild-type MBL. The B and C variants have disrupted Gly-Xaa_1-Xaa_2 repeats of the collagenous region resulting in an altered capacity to form the collagen triple helix. The D variant introduces an additional cysteine residue and so may disrupt oligomer formation by generation of additional disulphide bonds. Within the promoter region of the MBL gene nucleotide substitutions at positions −550, −221 and +4 give rise to H/L, Y/X and P/Q variants that influence the MBL expression. HY and LY haplotypes are associated with high, while LX haplotypes are associated with low, MBL plasma levels. Due to linkage disequilibrium, only seven haplotypes (HYPA, LYQA, LYPA, LXPA, LYPB, LYQC and HYPD) are commonly found. Determination of the genetic variability in MBL has been performed principally by M.W. Turner and

Figure 2 | Overview of the complement system

The complement cascade can be activated by three distinct pathways on recognition of a target. All three pathways lead ultimately to the formation of multicomponent serine proteases (C3 convertases) C4b2a and C3bBb, which allows these three routes to converge to the activation of C3. The covalent binding of this protein to activating surfaces is responsible for initiating most of the effects of the system. The role of ficolins is described in [49]. MAC, membrane attack complex.



P. Garred and their colleagues, and is described in recent reviews [26,27].

MBL and complement

MBL is the only member of the collectin family of proteins to activate the complement system. MBL can act directly as an opsonin, by binding to carbohydrates on pathogens, then interacting with MBL receptors on phagocytic cells (for further discussion of receptors, see [28]). However, it can also trigger the opsonic activity of complement, resulting in deposition of C3b/iC3b on targets, and stimulation of phagocytic uptake via the C3 receptors, CR1, CR3 and CR4.

The complement system is a major mediator of innate immune defence and contributes to many innate immune system functions including inflammation, opsonization and lysis. It consists of more than 30 proteins, both soluble in serum and bound to cell membranes. The complement system can be activated via three pathways (classical, lectin or alternative) upon recognition of PAMPs (Figure 2).

Recognition by the first two of these pathways leads to activation of a cascade of proteases with cleavage of the complement components C2 and C4 leading to the formation of the protease complex C4b2a, which cleaves C3 into C3a and C3b. C3b is a central component of the complement system due to its many binding interactions [29]. It binds covalently to complement-activating surfaces where it (and its breakdown product iC3b) act as opsonins for phagocytes, and C3b can also combine with a C3 convertase to form a C5 convertase. The C5 convertase initiates the assembly of the membrane attack complex (MAC) consisting of complement components C5–C9, which forms pores in the lipid bilayers of its targets, potentially causing cell lysis [30]. C3a acts as an anaphylotoxin mediating inflammatory responses along with C4a and C5a, other complement component fragments released during the complement cascade (Figure 2).

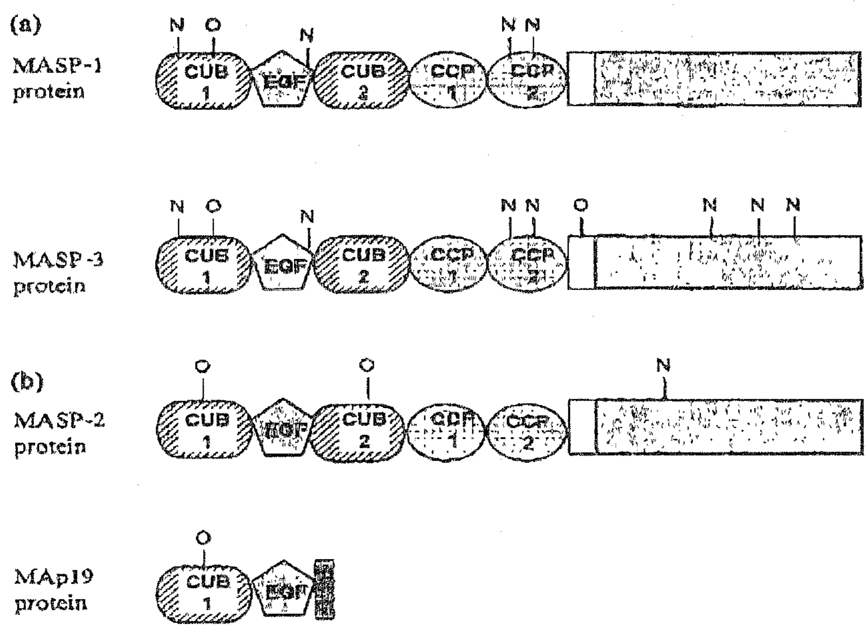
The pattern-recognition molecule of the classical pathway is C1q, found complexed with two protease proenzymes, C1r and C1s in the C1 complex ($C1qr_2s_2$) [31,32]. C1q binds to immunoglobulin G or M in immune complexes and, without need for immunoglobulin, to a wide range of charge patterns on target surfaces [33]. Upon binding of C1q to a target, C1r autoactivates and cleaves C1s to its active form. C1s then goes on to cleave C4, releasing C4a and a larger fragment, C4b. C4b binds to C2 and C1s cleaves the C2 to release C2b, leaving C2a complexed with C4b to form a C3 convertase (C4b2a).

C1q is similar in quaternary structure to MBL, and when MBL was first shown to activate complement, it was thought to act just like C1q, forming a complex with C1r and C1s [34]. This can indeed happen *in vitro*, but *in vivo* MBL instead interacts preferentially with homologues of C1r and C1s, called MBL-associated serine proteases (MASPs) [14,35,36].

The lectin pathway of complement is initiated through the binding of MBL or ficolins to carbohydrate structures present on a wide range of micro-organisms including bacteria, viruses, fungi and parasites. MBL can also recognize

Figure 3 | MASP domain organization

The MASPs are composed of an N-terminal complement-subcomponent C1r/C1s-like domain (CUB) domain, followed by an epidermal growth factor-like domain (EGF), a second CUB domain, two complement control protein domains (CCPs) and a C-terminal serine protease domain [45]. Three MASPs have been described so far, MASP-1, MASP-2 and more recently MASP-3 [14], MASP-3 is an alternative-splicing product from the MASP-1 gene, carrying the same heavy chain but a different serine protease domain. MAp19 consists of the first two domains of MASP-2 and a unique C-terminal sequence (EQSL). MASP-1, -2 and -3 are synthesized as proenzymes of 91, 78 and 101.5 kDa respectively. These molecular masses are calculated from their amino-acid sequences and by assuming that all the potential N-glycosylation sites are occupied by a carbohydrate of 3.5 kDa. Potential N- and O-glycosylation sites (N and O) are shown.



carbohydrate structures on antibodies, including the common IgG glycosylation variant IgG-G0 [37] and polymeric IgA [38]. MBL and ficolins are found in serum complexed with MASP proenzymes, and a small 19 kDa protein (MAp19). The MASPs are structurally similar to C1r and C1s with identical domain composition (Figure 3). The mechanisms for MASP activation have yet to be fully determined but upon MBL binding to a target, MASP-1 and MASP-2 are activated independently (unlike C1r and C1s), and MASP-2 cleaves C4 and C2 to form the C3 convertase C4b2a [39,40].

The alternative pathway is activated by a mechanism quite different from the classical and lectin pathways (for review, see [41]). It reacts to targets according to surface features of both charge and neutral sugar. It also has both antibody-dependent (IgG–immune complexes) and antibody-independent modes of activation.

MASPs

Unlike the C1 complex of the classical pathway, which appears to be homogeneous and of fixed stoichiometry ($C1qr_2s_2$), MBL complexes with MASPs are likely to be heterogeneous. When MBL–MASP complexes are captured on mannan-coated surfaces, MASP-1 is more abundant than MASP-2, and they are not present in a fixed ratio [42,43]. The additional components MASP-3 (an alternatively spliced product from the MASP-1 gene [14]) and MAp19 [45] are of low abundance. These MBL–MASP complexes are able to cleave C2 and C4 to form the C3 convertase, C4bC2a. This activity has been demonstrated to be a property of MASP-2 but not of MASP-1 [39]. MASP-1 and MASP-2 appear to function separately and there is evidence to suggest that they bind to MBL independently, and activate independently of each other [12,42,44].

MASP-1 and MASP-2 are synthesized as proenzymes of 91 and 78 kDa respectively [35,36] (Figure 3). They are activated through cleavage of a conserved Arg–Ile peptide bond in a linker region connecting the serine protease domain to the rest of the enzyme. This cleavage results in an A-chain (the first five domains and the linker region) linked by a disulphide bond to a B-chain, the serine protease domain. The MASPs have a trypsin-like (basic specificity) active site consisting of the catalytic residues His, Asp and Ser with Asp as the substrate-specificity residue at −6 from the active-site Ser.

Unlike the proteases C1s and C1r that form heterotetramers with the classical pathway component C1q ($C1r_2s_2$) [32], MASP-1 and MASP-2 do not appear to function as one unit. Dahl et al. [14] indicated that different MBL oligomers were present in serum with differing MASP compositions. They found MASP-1 and MAp19 associated with smaller MBL oligomers and MASP-2 and MASP-3 with larger oligomers. The significance of this finding with respect to their biological functions has yet to be determined. As noted above, MASP-2 cleaves and activates C4 and C2 [39,40], and so behaves like C1s. However there is no definitive natural substrate identified for MASP-1 or MASP-3 [39,40]. MASP-1 is not like C1r, as it does not seem to influence MASP-2 activation, while C1r activates C1s. MASP-1 was initially reported to cleave C3 directly [46,47] but this now seems unlikely. Studies by Wong et al. [42], Ambrus et al. [48] and Hajela et al. [40] have shown that cleavage of the haemolytically inactive form of C3 [C3(H_2O)] is easily detectable and occurs more than 100-fold faster than that of haemolytically active C3 ('live' C3). The observation of cleavage of live C3 is probably due to spontaneous hydrolysis of live C3 to form C3(H_2O), which is then cleaved [40]. Cleavage of C3(H_2O) is unlikely to have biological significance, as C3(H_2O) has very low abundance in the blood. Soluble human MASP-1 has a thrombin-like substrate specificity, cleaving the coagulation proteins Factor XIII (plasma transglutaminase) and fibrinogen [40]. These reactions are much faster than observed for C3 and may indicate that fibrinogen and Factor XIII are significant substrates *in vivo*. However, the substrates of MASP-1 and MASP-3 have yet to be fully explored.

References

1 Malhotra, R., Haurum, J., Thiel, S. and Sim, R.B. (1992) Eur. J. Immunol. **22**, 1437–1445
2 Kawasaki, T., Etoh, R. and Yamashina, I. (1978) Biochem. Biophys. Res. Commun. **81**, 1018–1024
3 Wild, J., Robinson, D. and Winchester, B. (1983) Biochem. J. **210**, 167–174
4 Colley, K.J., Beranek, M.C. and Baenziger, J.U. (1988) Biochem. J. **256**, 61–68
5 Ihara, I., Harada, Y., Ihara, S. and Kawakami, M. (1982) J. Immunol. **128**, 1256–1260
6 Sastry, K., Herman, G.A., Day, L., Deignan, E., Bruns, G., Morton, C.C. and Ezekowitz, R.A. (1989) J. Exp. Med. **170**, 1175–1189
7 Ezekowitz, R.A., Day, L.E. and Herman, G.A. (1988) J. Exp. Med. **167**, 1034–1046
8 Taylor, M.E., Brickell, P.M., Craig, R.K. and Summerfield, J.A. (1989) Biochem. J. **262**, 763–771
9 Drickamer, K., Dordal, M.S. and Reynolds, L. (1986) J. Biol. Chem. **261**, 6878–6887
10 Holmskov, U., Malhotra, R., Sim, R.B. and Jensenius, J.C. (1994) Immunol. Today **15**, 67–74
11 Lipscombe, R.J., Sumiya, M., Summerfield, J.A. and Turner, M.W. (1995) Immunology **85**, 660–667
12 Chen, C.B. and Wallis, R. (2001) J. Biol. Chem. **276**, 25894–25902
13 Lu, J.H., Thiel, S., Wiedemann, H., Timpl, R. and Reid, K.B. (1990) J. Immunol. **144**, 2287–2294
14 Dahl, M.R., Thiel, S., Matsushita, M., Fujita, T., Willis, A.C., Christensen, T., Vorup-Jensen, T. and Jensenius, J.C. (2001) Immunity **15**, 127–135
15 Wong, N. and Sim, R.B. (1997) Biochem. Soc. Trans. **25**, 415
16 Hickling, T.P., Malhotra, R. and Sim, R.B. (1998) Mol. Med. **4**, 266–275
17 Weis, W.I., Drickamer, K. and Hendrickson, W.A. (1992) Nature (London) **360**, 127–134
18 Malhotra, R. and Sim, R.B. (1995) Trends Microbiol. **3**, 240–244
19 Haurum, J.S., Thiel, S., Jones, I.M., Fischer, P.B., Laursen, S.B. and Jensenius, J.C. (1993) Aids **7**, 1307–1313
20 Neth, O., Jack, D.L., Dodds, A.W., Holzel, H., Klein, N.J. and Turner, M.W. (2000) Infect. Immun. **68**, 688–693
21 Jack, D.L., Dodds, A.W., Anwar, N., Ison, C.A., Law, A., Frosch, M., Turner, M.W. and Klein, N.J. (1998) J. Immunol. **160**, 1346–1353
22 Tabona, P., Mellor, A. and Summerfield, J.A. (1995) Immunology **85**, 153–159
23 Schelenz, S., Malhotra, R., Sim, R.B., Holmskov, U. and Bancroft, G.J. (1995) Infect. Immun. **63**, 3360–3366
24 Soothill, J.F. and Harvey, B.A. (1976) Arch. Dis. Child. **51**, 91–99
25 Super, M., Thiel, S., Lu, J., Levinsky, R.J. and Turner, M.W. (1989) Lancet **II**, 1236–1239
26 Turner, M.W. and Hamvas, R.M. (2000) Rev. Immunogenet. **2**, 305–322
27 Madsen, H.O., Satz, M.L., Hogh, B., Svejgaard, A. and Garred, P. (1998) J. Immunol. **161**, 3169–3175
28 Kojima, M., Presanis, J.S. and Sim, R.B. (2003) in Glycoimmunology **3** (Adv. Exp. Med. Biol. series), Kluwer/Plenum, in the press
29 Law, S.K. and Dodds, A.W. (1997) Protein Sci. **6**, 263–274
30 Bhakdi, S. and Tranum-Jensen, J. (1991) Immunol. Today **12**, 318–320; discussion, 321
31 Gigli, I., Porter, R.R. and Sim, R.B. (1976) Biochem. J. **157**, 541–548
32 Arlaud, G.J. and Colomb, M.G. (1987) Mol. Immunol. **24**, 317
33 Sim, R.B. and Malhotra, R. (1994) Biochem. Soc. Trans. **22**, 106–111
34 Ohta, M., Okada, M., Yamashina, I. and Kawasaki, T. (1990) J. Biol. Chem. **265**, 1980–1984
35 Sato, T., Endo, Y., Matsushita, M. and Fujita, T. (1994) Int. Immunol. **6**, 665–669
36 Thiel, S., Vorup-Jensen, T., Stover, C.M., Schwaeble, W., Laursen, S.B., Poulsen, K., Willis, A.C., Eggleton, P., Hansen, S., Holmskov, U. et al. (1997) Nature (London) **386**, 506–510
37 Malhotra, R., Wormald, M.R., Rudd, P.M., Fischer, P.B., Dwek, R.A. and Sim, R.B. (1995) Nat. Med. **1**, 237–243
38 Roos, A., Bouwman, L.H., van Gijlswijk-Janssen, D.J., Faber-Krol, M.C., Stahl, G.L. and Daha, M.R. (2001) J. Immunol. **167**, 2861–2868
39 Vorup-Jensen, T., Petersen, S.V., Hansen, A.G., Poulsen, K., Schwaeble, W., Sim, R.B., Reid, K.B., Davis, S.J., Thiel, S. and Jensenius, J.C. (2000) J. Immunol. **165**, 2093–2100
40 Hajela, K., Kojima, M., Ambrus, G., Wong, K.H., Moffatt, B.E., Ferluga, J., Hajela, S., Gal, P. and Sim, R.B. (2002) Immunobiology **205**, 467–475
41 Sim, R.B. and Dodds, A.W. (1997) in Complement, a Practical Approach (Dodds, A.W. and Sim, R.B., eds.), pp. 1–18, Oxford University Press, Oxford
42 Wong, N.K., Kojima, M., Dobo, J., Ambrus, G. and Sim, R.B. (1999) Mol. Immunol. **36**, 853–861
43 Mayilyan, K.R., Presanis, J.S., Arnold, J.N., Hajela, K. and Sim, R.B. (2003) www.biochemsoctrans.org/bst/031/0315A013000B.pdf
44 Thielens, N.M., Cseh, S., Thiel, S., Vorup-Jensen, T., Rossi, V., Jensenius, J.C. and Arlaud, G.J. (2001) J. Immunol. **166**, 5068–5077
45 Stover, C.M., Thiel, S., Thelen, M., Lynch, N.J., Vorup-Jensen, T., Jensenius, J.C. and Schwaeble, W.J. (1999) J. Immunol. **162**, 3481–3490
46 Ogata, R.T., Low, P.J. and Kawakami, M. (1995) J. Immunol. **154**, 2351–2357
47 Matsushita, M. and Fujita, T. (1995) Immunobiology **194**, 443–448
48 Ambrus, G., Gal, P., Kojima, M., Szilagyi, K., Balczer, J., Antal, J., Graf, L., Laich, A., Moffatt, B.E., Schwaeble, W. et al. (2003) J. Immunol. **170**, 1374–1382
49 Matsushita, M., Endo, Y. and Fujita, T. (2000) J. Immunol. **164**, 2281–2284

Received 18 March 2003